Parallax Geometry of Pairs of Points for 3D Scene Analysis

Michal Irani and P. Anandan

David Sarnoff Research Center, CN5300, Princeton, NJ 08543-5300, USA

Abstract. We present a geometric relationship between the image motion of *pairs* of points over multiple frames. This relationship is based on the *parallax* displacements of points with respect to an arbitrary planar surface, and does not involve epipolar geometry. A constraint is derived over two frames for any pair of points, relating their projective structure (with respect to the plane) based only on their image coordinates and their parallax displacements. Similarly, a $3D$-rigidity constraint between pairs of points over multiple frames is derived. We show applications of these parallax-based constraints to solving three important problems in $3D$ scene analysis: (i) the recovery of $3D$ scene structure, (ii) the detection of moving objects in the presence of camera induced motion, and (iii) the synthesis of new camera views based on a given set of views. Moreover, we show that this approach can handle difficult situations for $3D$ scene analysis, e.g., where there is only a small set of parallax vectors, and in the presence of independently moving objects.

1 Introduction

The analysis of three dimensional scenes from image sequences has a number of goals. These include (but are not limited to): (i) the recovery of 3D scene structure, (ii) the detection of moving objects in the presence of camera induced motion, and (iii) the synthesis of new camera views based on a given set of views. The traditional approach to these types of problems has been to first recover the epipolar geometry between pairs of frames and then apply that information to achieve the abovementioned goals. However, this approach is plagued with the difficulties associated with recovering the epipolar geometry [24].

Recent approaches to 3D scene analysis have overcome some of the difficulties in recovering the epipolar geometry by decomposing the motion into a combination of a planar homography and residual parallax [11, 15, 17]. However, they still require the explicit estimation of the epipole itself, which can be difficult in many cases.

More recently, progress has been made towards deriving constraints directly based on collections of points in multiple views. Examples are the trilinearity constraints [18, 16] which eliminate the scene structure in favor of the camera geometries; the dual-shape tensor [23] which eliminates the camera motion in favor of scene structure; the more general framework of multipoint multiview geometry [6, 3]; and the work on multiple view invariants without requiring the recovery of the epipolar geometry [24]. In its current form, this class of methods

does not address the problem of shape recovery in *dynamic* scenes, in particular when the amount of image motion due to independent moving object is not negligible.

In this paper we develop geometric relationships between the residual (planar) parallax displacements of *pairs* of points. These geometric relationships address the problem of 3D scene analysis even in *difficult* conditions, i.e., when the epipole estimation is ill-conditioned, when there is a small number of parallax vectors, and in the presence of moving objects. We show how these relationships can be applied to each of the three problems outlined at the beginning of this section. Moreover, the use of the parallax constraints derived here provides a continuum between "$2D$ algorithms" and the "$3D$ algorithms" for each of the problems mentioned above.

In Section 2 a *parallax-based structure constraint* is derived, which relates the *projective structure* of two points to their image positions and their *parallax displacements* alone. By eliminating the relative projective structure of a pair of points between *three* frames, we arrive at a constraint on the parallax displacements of two points *moving as a rigid object* over those frames. We refer to this as the *parallax-based rigidity constraint.*

In Section 3 an alternative way of deriving the *parallax-based rigidity constraint* is presented, this time geometrically rather than algebraically. This leads to a simple and intuitive geometric interpretation of the multiframe rigidity constraint and to the derivation of a *dual point* to the epipole.

In Section 4 the pairwise parallax-based constraints are applied to solving three important problems in $3D$ scene analysis, even in the abovementioned difficult scenarios: (i) the recovery of $3D$ scene structure, (ii) the detection of moving objects in the presence of camera induced motion, and (iii) the synthesis of new camera views based on a given set of views.

2 Parallax-Based Constraints on Pairs of Points

In this section we derive a constraint on the parallax motion of pairs of points between two frames. We show how this constraint can be used to recover relative $3D$ structure of two points from their parallax vectors alone, without any additional information, and in particular, without requiring the recovery of the camera epipoles. The parallax constraint is then extended to *multiple* frames and to multiple points to obtain *rigidity* constraints on image points based on their parallax displacements alone, without involving any scene or camera geometry.

2.1 The Planar Parallax Decomposition

To derive the parallax constraint, we first briefly describe the decomposition of the image motion into a homography (i.e., the image motion of an arbitrary planar surface) and residual parallax displacements. This decomposition has been previously derived and used in [11, 15, 17]. For a more detailed derivation see also [7].

Fig. 1 provides a geometric interpretation of the planar parallax. Let $\mathbf{P} = (X, Y, Z)^T$ and $\mathbf{P'} = (X', Y', Z')^T$ denote the Cartesian coordinates of a scene point with respect to two different camera views, respectively. Let $\mathbf{p} = (x, y)^T$

and $\mathbf{p}' = (x', y')^T$ respectively denote the corresponding coordinates of the corresponding image points in the two image frames. Let $\mathbf{T} = (T_x, T_y, T_z)$ denote

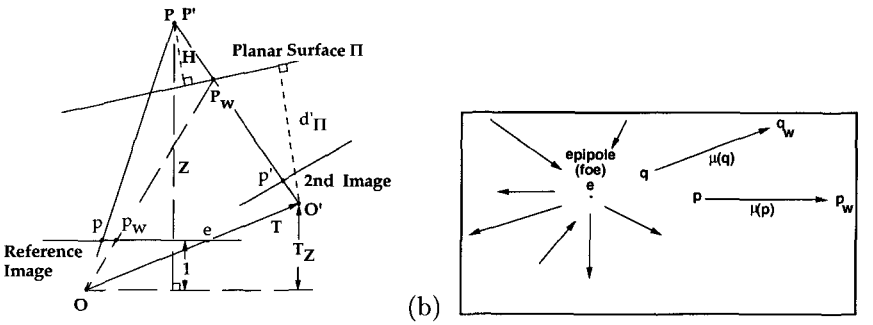

Fig. 1. *The plane+parallax decomposition. (a) The geometric interpretation. (b) The epipolar field of the residual parallax displacements.*

the camera translation between the two views. Let Π denote an arbitrary (*real* or *virtual*) planar surface in the scene, and let A' denote the homography that aligns the planar surface Π between the second and first frame (i.e., for all points $\mathbf{P} \in \Pi$, $\mathbf{P} = A'\mathbf{P}'$). It can be shown (see [11, 9, 15, 17]) that the 2D image displacement of the point $\mathbf{P}$ can be written as $\mathbf{u} = (\mathbf{p}' - \mathbf{p}) = \mathbf{u}_\pi + \boldsymbol{\mu}$, where $\mathbf{u}_\pi$ denotes the *planar* part of the $2D$ image motion (the homography due to Π), and $\boldsymbol{\mu}$ denotes the residual *planar parallax* $2D$ motion. The homography due to Π results in an image motion field that can be modeled as a 2D *projective transformation*. When $T_z \neq 0$:

$$\mathbf{u}_\pi = (\mathbf{p}' - \mathbf{p_w}) \; ; \; \boldsymbol{\mu} = \gamma \frac{T_z}{d'_\pi}(\mathbf{e} - \mathbf{p_w}) \tag{1}$$

where $\mathbf{p_w}$ denotes the image point in the first frame which results from warping the corresponding point $\mathbf{p}'$ in the second image by the 2D parametric transformation of the plane Π. The $2D$ image coordinates of the epipole (or the *focus-of-expansion*, FOE) in the first frame are denoted by $\mathbf{e}$, and d'_π is the perpendicular distance from the second camera center to the reference plane (see Fig. 1). γ is a measure of the 3D shape of the point $\mathbf{P}$. In particular, $\gamma = \frac{H}{Z}$, where H is the perpendicular distance from the $\mathbf{P}$ to the reference plane, and Z is the "range" (or "depth") of the point $\mathbf{P}$ with respect to the first camera. We refer to γ as the projective 3D structure of the point $\mathbf{P}$. In the case when $T_z = 0$, the parallax motion μ has a slightly different form: $\mu = \frac{\gamma}{d'_\pi}\mathbf{t}$, where $t = (T_x, T_y)^T$.

2.2 The Parallax Based Structure Constraint

Theorem 1: Given the planar-parallax displacement vectors $\boldsymbol{\mu_1}$ and $\boldsymbol{\mu_2}$ of two points that belong to the static background scene, their *relative 3D projective*

structure $\frac{\gamma_2}{\gamma_1}$ is given by:

$$\frac{\gamma_2}{\gamma_1} = \frac{\boldsymbol{\mu_2}^T(\Delta \mathbf{p_w})_\perp}{\boldsymbol{\mu_1}^T(\Delta \mathbf{p_w})_\perp}, \tag{2}$$

where, as shown in Fig. 2.a, $\mathbf{p_1}$ and $\mathbf{p_2}$ are the image locations (in the reference frame) of two points that are part of the static scene, $\Delta \mathbf{p_w} = \mathbf{p_{w_2}} - \mathbf{p_{w_1}}$, the vector connecting the "warped" locations of the corresponding second frame points (as in Eq. (1)), and $\mathbf{v}_\perp$ signifies a vector perpendicular to $\mathbf{v}$.

Proof: From Eq. (1), we know that $\boldsymbol{\mu_1} = \gamma_1 \frac{T_z}{d'_\pi}(\mathbf{e} - \mathbf{p_{w_1}})$ and $\boldsymbol{\mu_2} = \gamma_2 \frac{T_z}{d'_\pi}(\mathbf{e} - \mathbf{p_{w_2}})$. Therefore,

$$\boldsymbol{\mu_1}\gamma_2 - \boldsymbol{\mu_2}\gamma_1 = \gamma_1\gamma_2 \frac{T_Z}{d'}(\mathbf{p_{w_2}} - \mathbf{p_{w_1}}) \tag{3}$$

This last step eliminated the epipole $\mathbf{e}$. Eq. (3) entails that the vectors on both sides of the equation are parallel. Since $\gamma_1\gamma_2\frac{T_Z}{d'}$ is a scalar, we get:

$$(\boldsymbol{\mu_1}\gamma_2 - \boldsymbol{\mu_2}\gamma_1) \parallel \Delta \mathbf{p_w} \Rightarrow (\boldsymbol{\mu_1}\gamma_2 - \boldsymbol{\mu_2}\gamma_1)^T(\Delta \mathbf{p_w})_\perp = 0 \Rightarrow \frac{\gamma_2}{\gamma_1} = \frac{\boldsymbol{\mu_2}^T(\Delta \mathbf{p_w})_\perp}{\boldsymbol{\mu_1}^T(\Delta \mathbf{p_w})_\perp}, \tag{4}$$

which is the *pairwise parallax constraint*. When $T_Z = 0$, a constraint stronger than Eq. (4) can be derived: $(\boldsymbol{\mu_1}\frac{\gamma_2}{\gamma_1} - \boldsymbol{\mu_2}) = 0$. However, Eq. (4), still holds. This is important, as we do not have a-priori knowledge of T_Z to distinguish between the two cases. ∎

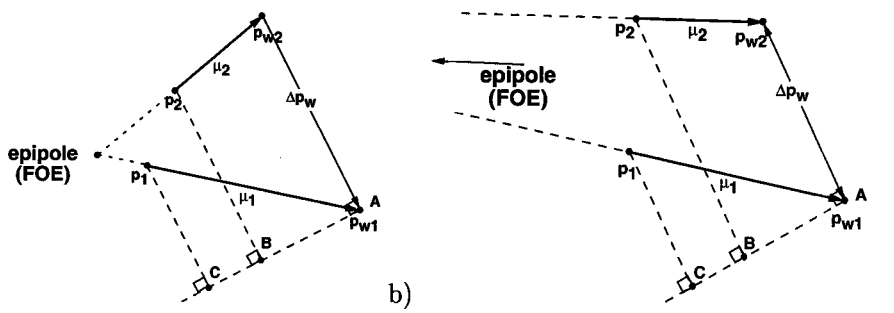

Fig. 2. The relative structure constraint. *(a) This figure geometrically illustrates the relative structure constraint (Eq. 2):* $\frac{\gamma_2}{\gamma_1} = \frac{\boldsymbol{\mu_2}^T(\Delta \mathbf{p_w})_\perp}{\boldsymbol{\mu_1}^T(\Delta \mathbf{p_w})_\perp} = \frac{\mathbf{AB}}{\mathbf{AC}}$. *(b) When the parallax vectors are nearly parallel, the epipole estimation is unreliable. However, the relative structure* $\frac{\mathbf{AB}}{\mathbf{AC}}$ *can be reliably computed even in this case.*

Fig. 2.a displays the constraint geometrically. The fact that relative structure of one point with respect to another can be obtained using only the two parallax vectors is not surprising: *In principle*, one could use the two parallax vectors to recover the epipole (the intersection point of the two vectors), and then use the magnitudes and distances of the points from the computed epipole to estimate their relative projective structure. The benefit of the constraint (2) is that it

provides this information *directly* from the positions and parallax vectors of the two points, without the need to go through the computation of the epipole, using as much information as one point can give on another. Fig. 2.b graphically shows an example of a configuration in which estimating the epipole is very unreliable, whereas estimating the relative structure *directly* from Eq. (2) *is* reliable.

2.3 The Parallax-Based Rigidity Constraint

In this section we extend the parallax-based structure constraint to *multiple frames* and to *multiple points*, to obtain *rigidity* constraints that are based only on parallax displacements of the image points, and involve neither *structure* parameters nor *camera geometry*.

Rigidity Over Multiple Frames: Let $\mathbf{p_1}$ and $\mathbf{p_2}$ be two *image* points in the first (reference) frame. Let $\boldsymbol{\mu_1^j}, \boldsymbol{\mu_2^j}$ be the parallax displacement displacements of the two points between the reference frame and the jth frame, and $\boldsymbol{\mu_1^k}, \boldsymbol{\mu_2^k}$ be the parallax displacements between the reference frame and the kth frame. Let $(\Delta\mathbf{p_w})^j, (\Delta\mathbf{p_w})^k$ be the corresponding vectors connecting the warped points as in Eq. (2) and Fig. 2. Using the relative structure constraint (2), for any two frames j and k we get: $\frac{\gamma_2}{\gamma_1} = \frac{\boldsymbol{\mu_2}^{\mathbf{j}^T}(\Delta\mathbf{p_w})^{\mathbf{j}}_{\perp}}{\boldsymbol{\mu_1}^{\mathbf{j}^T}(\Delta\mathbf{p_w})^{\mathbf{j}}_{\perp}} = \frac{\boldsymbol{\mu_2}^{\mathbf{k}^T}(\Delta\mathbf{p_w})^{\mathbf{k}}_{\perp}}{\boldsymbol{\mu_1}^{\mathbf{k}^T}(\Delta\mathbf{p_w})^{\mathbf{k}}_{\perp}}$.

Multiplying by the denominators yields the *rigidity constraint* of the two points over three frames (reference frame, frame j, and frame k):

$$(\boldsymbol{\mu_1}^{\mathbf{k}^T}(\Delta\mathbf{p_w})^{\mathbf{k}}_{\perp})(\boldsymbol{\mu_2}^{\mathbf{j}^T}(\Delta\mathbf{p_w})^{\mathbf{j}}_{\perp}) - (\boldsymbol{\mu_1}^{\mathbf{j}^T}(\Delta\mathbf{p_w})^{\mathbf{j}}_{\perp})(\boldsymbol{\mu_2}^{\mathbf{k}^T}(\Delta\mathbf{p_w})^{\mathbf{k}}_{\perp}) = 0. \qquad (5)$$

Thus, the planar parallax motion trajectory of a single image point (e.g., $\mathbf{p_1}$) over several frames constrains the planar parallax motion trajectory of any other point (e.g., $\mathbf{p_2}$) according to Eq. (5). The rigidity constraint (5) can therefore be applied to detect inconsistencies in the $3D$ motion of two image points (i.e., say whether the two image points are projections of $3D$ points belonging to a same or different $3D$ moving objects) based on their *parallax* motion among three (or more) frames alone, without the need to estimate either *camera geometry* or *structure* parameters. In contrast to previous approaches (e.g., the trilinear tensor [16]), when planar parallax motion is available, Eq. (5) provides certain advantages: (i) it is based on the parallax motion of a *single* image point, (ii) it does not require any numerical estimation (e.g., unlike [16], it does not require estimation of tensor parameters), (iii) it does not involve, *explicitly or implicitly*, any shape or camera geometry information other than that already implicit in the planar parallax motion itself.

Rigidity Over Multiple Points: Instead of considering pairs of points over multiple frames, we can consider multiple points over pairs of frames to obtain a different form of the rigidity constraint.

Let $\mathbf{p_1}$, $\mathbf{p_2}$, and $\mathbf{p_3}$ be three image points in the first (reference) frame. Let $\boldsymbol{\mu_i}$ denote the $2D$ planar parallax motion of $\mathbf{p_i}$ from the first frame to another frame ($i = 1, 2, 3$).

Using the shape invariance constraint (2):

$$\frac{\gamma_2}{\gamma_1} = \frac{\boldsymbol{\mu_2}^T(\Delta \mathbf{p_{w_{2,1}}})_\perp}{\boldsymbol{\mu_1}^T(\Delta \mathbf{p_{w_{2,1}}})_\perp} \; ; \; \frac{\gamma_3}{\gamma_2} = \frac{\boldsymbol{\mu_3}^T(\Delta \mathbf{p_{w_{3,2}}})_\perp}{\boldsymbol{\mu_2}^T(\Delta \mathbf{p_{w_{3,2}}})_\perp} \; ; \; \frac{\gamma_3}{\gamma_1} = \frac{\boldsymbol{\mu_3}^T(\Delta \mathbf{p_{w_{3,1}}})_\perp}{\boldsymbol{\mu_1}^T(\Delta \mathbf{p_{w_{3,1}}})_\perp}.$$

Equating $\frac{\gamma_3}{\gamma_1} = \frac{\gamma_3}{\gamma_2}\frac{\gamma_2}{\gamma_1}$, and multiplying by the denominators, we get the rigidity constraint for three points over a pair of frames:

$$(\boldsymbol{\mu_3}^T \Delta \mathbf{p_{w_{3,2}}}_\perp)(\boldsymbol{\mu_2}^T \Delta \mathbf{p_{w_{2,1}}}_\perp)(\boldsymbol{\mu_1}^T \Delta \mathbf{p_{w_{3,1}}}_\perp) = (\boldsymbol{\mu_2}^T \Delta \mathbf{p_{w_{3,2}}}_\perp)(\boldsymbol{\mu_1}^T \Delta \mathbf{p_{w_{2,1}}}_\perp)(\boldsymbol{\mu_3}^T \Delta \mathbf{p_{w_{3,1}}}_\perp). \tag{6}$$

The fact that three points in two frames form a rigidity constraint is not surprising: *In principle*, one could use two of the three parallax vectors to obtain the epipole (the intersection point of the two vectors). $3D$ rigidity will constrain the parallax vector of the third point to lie on the epipolar line emerging from the computed epipole through the third point. The benefit of the rigidity constraint (6) is in the fact that it provides this information directly from the positions and parallax vectors of the three points, without the need to go through the unstable computation of the epipole, using as much information as two point can give on the third.

2.4 The Generalized Parallax Constraint

The pairwise-parallax constraint (Eq. (2)) can be extended to handle full image motion (as opposed to *parallax* motion), even when the homography is unknown. Eq. (1) can be rewritten (in homogeneous coordinates) as [11, 15, 17]:

$$\mathbf{p} = \mathbf{p_w} + \gamma \frac{1}{d'_\pi}(T_Z \mathbf{p_w} - \mathbf{T}) = \frac{A'\mathbf{p'}}{a'^T_3 \mathbf{p'}} + \gamma \frac{1}{d'_\pi}(T_Z \frac{A'\mathbf{p'}}{a'^T_3 \mathbf{p'}} - \mathbf{T}), \tag{7}$$

where A' is an *unknown* homography from frame2 to the first frame, and $\mathbf{a'_3}$ is the third row of the 3×3 matrix $\mathbf{A'}$. It could relate to *any* planar surface in the scene, in particular a *virtual* plane.

Given two points $\mathbf{p_1}$ and $\mathbf{p_2}$, we can eliminate $\mathbf{T}$ in a manner similar to that done in Theorem 1. This yields the *generalized parallax constraint* in terms the relative projective structure $\frac{\gamma_2}{\gamma_1}$:

$$(\frac{\gamma_2}{\gamma_1}(\mathbf{p_1} - \frac{A'\mathbf{p'}_1}{a'^T_3 \mathbf{p'}_1}) - (\mathbf{p_2} - \frac{A'\mathbf{p'}_2}{a'^T_3 \mathbf{p'}_2}))^T (\frac{A'\mathbf{p'}_1}{a'^T_3 \mathbf{p'}_1} - \frac{A'\mathbf{p'}_2}{a'^T_3 \mathbf{p'}_2})_\perp = 0. \tag{8}$$

The generalized parallax constraint suggests a new implicit representation of *general* $2D$ image motion: Rather than looking at the representation of $2D$ image motion in terms of: *homography + epipole + projective structure* [11, 15, 17] it suggests an implicit representation of $2D$ image motion in terms of: *homography + relative projective structure* of pairs of points. Since this representation does not contain the epipole, it can be easily extended to multiple frames. In a similar manner, the *rigidity constraints* (5) and (6) can also be generalized to handle full image motion with unknown homography. For more details see [7].

3 Parallax Geometry and an Epipole Dual

In this section, we present a geometric view of the parallax-based rigidity constraint. This leads to derivation of a *dual point* to the epipole.

The 3D geometric structure associated with the planar parallax of pairs of points between two frames is illustrated in Fig. 3.a. In this figure, Π is the planar surface, and $\mathbf{P}$ and $\mathbf{Q}$ are the two scene points. As in the case of Fig. 1, $\mathbf{P_w}$ and $\mathbf{Q_w}$ are the intersections of rays $\mathbf{O'P}$ and $\mathbf{O'Q}$ with the plane Π. The points $\mathbf{p_w}$ and $\mathbf{q_w}$ on the reference image are the projections of $\mathbf{P_w}$ and $\mathbf{Q_w}$, and are therefore the points to which the planar homography transforms $\mathbf{p'}$ and $\mathbf{q'}$ respectively. Below, we refer to $\mathbf{p_w}$ and $\mathbf{q_w}$ as "warped points".

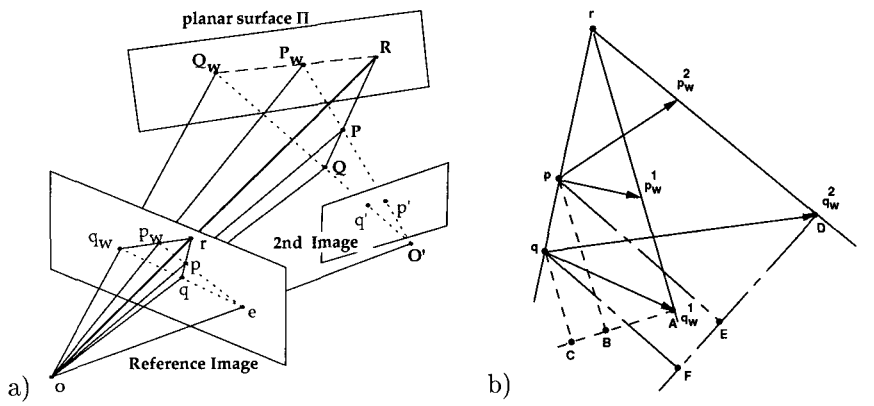

Fig. 3. The dual of the epipole. *(a) The line connecting points* $\mathbf{p}, \mathbf{q}$ *in the reference image intersects the line connecting the warped points* $\mathbf{p_w}, \mathbf{q_w}$ *at* $\mathbf{r}$. *The point* $\mathbf{r}$ *is the image of the point* $\mathbf{R}$ *which is the intersection of the line* $\mathbf{PQ}$ *on the planar surface* Π. *(b) The line connecting points* $\mathbf{p}, \mathbf{q}$ *in the reference image and the lines connecting the corresponding warped points* $\mathbf{fp_w^1}, \mathbf{q_w^1}$ *and* $\mathbf{p_w^2}, \mathbf{q_w^2}$ *from other frames all intersect at* $\mathbf{r}$, *the dual of the epipole.*

Let $\mathbf{R}$ be the intersection of the line connecting $\mathbf{P}$ and $\mathbf{Q}$ with the plane Π, and $\mathbf{r}$ be its projection on the reference image plane. Since $\mathbf{P}, \mathbf{Q}, \mathbf{R}$ are colinear and $\mathbf{P_w}, \mathbf{Q_w}, \mathbf{R}$ are colinear, therefore $\mathbf{p_w}, \mathbf{q_w}, \mathbf{r}$ and colinear and $\mathbf{p}, \mathbf{q}, \mathbf{r}$ are colinear. *In other words, the line connecting* $\mathbf{p_w}$ *and* $\mathbf{q_w}$ *and the line connecting* $\mathbf{p}$ *and* $\mathbf{q}$ *intersect at* $\mathbf{r}$, *the image of the point* $\mathbf{R}$. (See also [24] for the same observation.)

Note that the point $\mathbf{r}$ does not depend on the second camera view. Therefore, if multiple views are considered, then the lines connecting the warped points $\mathbf{p_w^j}$ and $\mathbf{q_w^j}$ (for any frame j), meet at $\mathbf{r}$ for all such views.

The convergence of the lines is illustrated in Fig. 3.b. Referring to that figure, since the lines $\mathbf{qC}, \mathbf{pB}$ and $\mathbf{rA}$ are parallel to each other and intersect the lines $\mathbf{qpr}$ and $\mathbf{CAB}$: $\frac{\mathbf{qr}}{\mathbf{pr}} = \frac{\mathbf{CA}}{\mathbf{BA}} = \frac{\gamma_q}{\gamma_p}$. Similarly, $\frac{\mathbf{qr}}{\mathbf{pr}} = \frac{\mathbf{FD}}{\mathbf{ED}} = \frac{\gamma_q}{\gamma_p}$. Hence $\frac{\mathbf{qr}}{\mathbf{pr}} = \frac{\mathbf{CA}}{\mathbf{BA}} = \frac{\mathbf{FD}}{\mathbf{ED}}$. This is the same as the rigidity constraint of a pair of points over multiple frames derived in Section 2. Note, however, the rigidity constraint itself does not

require the estimation of the point of convergence $\mathbf{r}$, just as it does not require the estimation of the epipole.

The point $\mathbf{r}$ is the dual of the epipole: the epipole is the point of intersection of multiple parallax vectors between a pair of frames, i.e., the point of intersection of all lines connecting each image point with its warped point between a pair of frames. Whereas the dual point $\mathbf{r}$ is the point of intersection of all lines connecting a pair of points in the reference image and the corresponding pair of warped points from all other frames.

4 Applications of Pairwise Parallax Geometry

In this section we show how parallax geometry in its various forms, which was introduced in the previous sections, provides an approach to handling some well-known problems in $3D$ scene analysis, in particular: (i) Moving object detection, (ii) Shape recovery, (ii) New view generation.

An extensive literature exists on methods for solving the above mentioned problems. They can be roughly classified into two main categories: (i) *2D methods* (e.g.,[8, 2, 13]): These methods assume that the image motion of the scene can be described using a $2D$ parametric transformation. They handle *dynamic* scenarios, but are limited to planar scenes or to very small camera translations. These fail in the presence of parallax motion. (ii) *3D methods* (e.g., [11, 15, 17, 5, 14, 23, 6, 3]): These methods handle general $3D$ scenes, but are (in their current form) limited to static scenarios or to scenarios where the parallax is both dense and of significant magnitude in order to overcome "noise" due to moving objects.

The use of the parallax constraints derived here provides a continuum between "$2D$ algorithms" and the "$3D$ algorithms". The need for bridging this gap exists in realistic image sequences, because it is not possible to predict in advance which situation would occur. Moreover, both types of scenarios can occur within the same sequence, with gradual transitions between them.

Estimating Planar Parallax Motion: The estimation of the planar parallax motion used for performing the experiments presented in this section was done using two successive computational steps: (i) $2D$ image alignment to compensate for a detected planar motion (i.e., the homography) in the form of a $2D$ parametric transformation, and, (ii) estimation of residual image displacements between the aligned images (i.e, the parallax).

We use previously developed methods [1, 8] in order to compute the $2D$ *parametric* image motion of a single $3D$ planar surface in the scene. These techniques lock onto a "dominant" planar motion in an image pair, even in the presence of other differently moving objects in the field of view. The estimated parametric motion is used to warp the second image frame to the first. The *residual* image displacements (e.g., parallax vectors) are then estimated using the optical flow estimation technique described in [1].

4.1 Moving Object Detection

A number of techniques exist to handle multiple motions analysis in the simpler $2D$ case, where motions of independent moving objects are modeled by $2D$ parametric transformation [8, 2, 13]. These methods, however, would also detect

points with planar parallax motion as moving objects, as they have a different $2D$ image motion than the planar part of the background scene.

In the general $3D$ case, the moving object detection problem is much more complex, since it requires detecting $3D$ motion inconsistencies. Typically, this is done by recovering the epipolar geometry. Trying to estimate epipolar geometry (i.e., camera motion) in the presence of multiple moving objects, with no prior segmentation, is extremely difficult. This problem becomes even more acute when there exists only sparse parallax information. A careful treatment of the issues and problems associated with moving object detection in 3D scenes is given in [19]. Methods have been proposed for recovering camera geometry in the presence of moving objects [12, 21] in cases when the available parallax was *dense* enough and the independent motion was *sparse* enough to be treated as *noise*.

Fig. 4.a graphically displays an example of a configuration in which estimating the epipole in presence of multiple moving objects can be very erroneous, even when using clustering techniques in the epipole domain as suggested by [12, 20]. Relying on the epipole computation to detect inconsistencies in $3D$ motion fails in detecting moving objects in such cases.

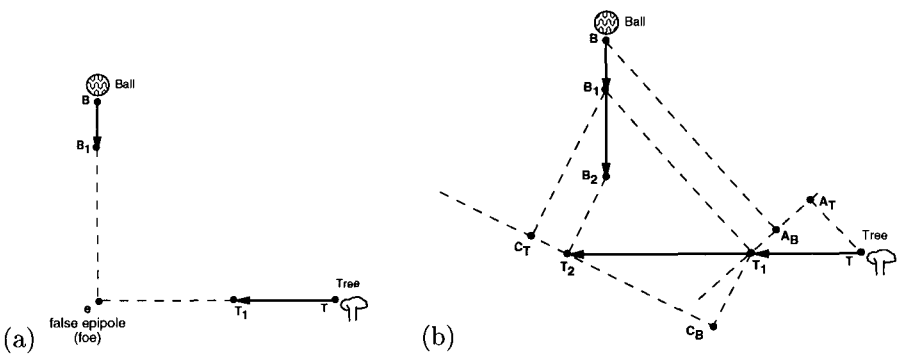

Fig. 4. Reliable detection of 3D motion inconsistency with sparse parallax information. *(a) Camera is translating to the right. The only static object with pure parallax motion is that of the tree. Ball is falling independently. The* incorrectly *estimated epipole* e *is consistent with both motions. (b) The rigidity constraint applied to this scenario detects 3D inconsistency over three frames, since* $\frac{\mathbf{T_1A_B}}{\mathbf{T_1A_T}} \neq \frac{\mathbf{T_2C_B}}{-\mathbf{T_2C_T}}$. *In this case, even the signs do not match.*

The parallax rigidity constraint(Eq. (5)) can be applied to detect inconsistencies in the $3D$ motion of one image point relative to another directly from their "parallax" vectors over multiple (three or more) frames, without the need to estimate either *camera geometry* or *shape* parameters. This provides a useful mechanism for clustering (or segmenting) the "parallax" vectors (i.e., the residual motion after planar registration) into consistent groups belonging to consistently $3D$ moving objects, even in cases such as in Fig. 4.a, where the parallax information is minimal, and the independent motion is not negligible. Fig. 4.b graphically explains how the rigidity constraint (5) detects the $3D$ inconsistency of Fig. 4.a over three frames.

Fig. 5 shows an example of using the rigidity constraint (5) to detect $3D$ inconsistencies.

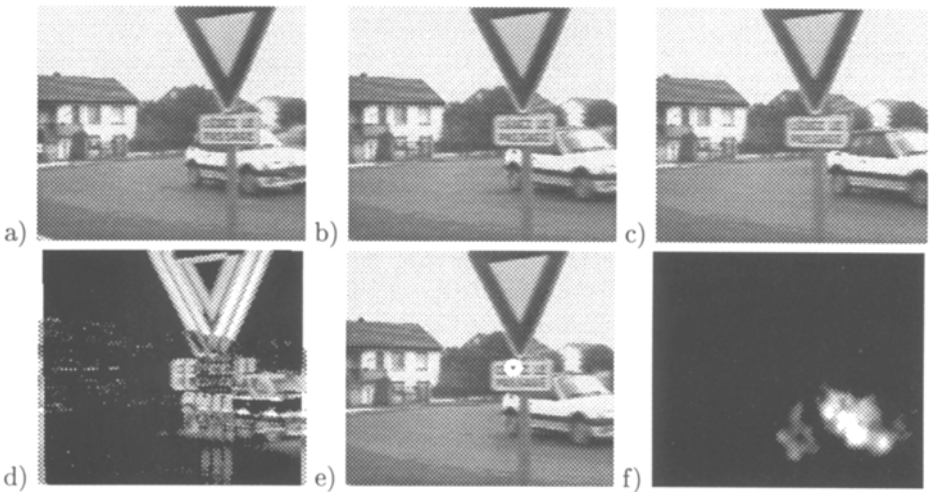

Fig. 5. Moving object detection relying on a single parallax vector.
(a,b,c) Three image frames from a sequence obtained by a camera translating from left to right, inducing parallax motion of different magnitudes on the house, road, and road-sign. The car moves independently from left to right. The middle frame (Fig. 5.b) was chosen as the frame of reference. (d) Differences taken after 2D image registration. The detected 2D planar motion was that of the house, and is canceled by the 2D registration. All other scene parts that have different 2D motions (i.e., parallax *motion or* independent *motion) are misregistered. (e) The selected point of reference (on the road-sign) highlighted by a white circle. (f) The measure of 3D-inconsistency of all points in the image with respect to the road-sign point. Bright regions indicate violations in 3D rigidity detected over three frames. These regions correspond to the car. Regions close to the image boundary were ignored. All other regions of the image appear to move 3D-consistently with the road-sign point.*

In [16] a rigidity constraint between three frames in the form of the trilinear tensor has been presented using regular image displacements. However, it requires having a collection of a set of image points which is known *a priori* to belong to the single $3D$ moving object. Selecting an inconsistent set of points will lead to an erroneous tensor. In [22] the trilinear constraint was used to segment and group multiple moving objects using robust techniques.

4.2 Shape Recovery

Numerous methods for recovering 3D depth from multiple views of calibrated cameras have been suggested. More recently, methods have been developed for recovering *projective structure* or *affine structure* from *uncalibrated* cameras [5, 14]. Under this category, methods for recovering structure using *planar parallax* motion have been proposed [10, 11, 15, 17]. Those methods rely on prior estimation

of partial or full camera geometry. In particular, they rely on estimating the camera epipole. Some methods for recovering camera or scene geometry in the presence of very sparse independent motion have been suggested [12, 20].

Recently, a complete theory has been presented [23, 3] for estimating the shape *directly* from image displacements over multiple frames, *without* the need to recover the epipolar geometry. These assume, however, that the scene is static. The problem of shape recovery in dynamic scenes, where the amount of image motion due to independent moving object is not negligible, is not addressed in these works.

The parallax-based structure constraint (Eq. (2)) can be used to recover a the 3D relative structure between pairs of points directly from their parallax vectors. This implies that the structure of the entire scene can be recovered relative to a *single* reference image point (with non-zero parallax). Singularities occur when the denominator of constraint (Eq. (2)) tends to zero, i.e., for points that lie on the line passing through the reference point in the direction of its parallax displacement vector.

Fig. 6 shows an example of recovering structure of an entire scene relative to a *single* reference point. Fig. 6.f shows the recovered relative structure of the entire scene from *two* frames (Figs. 6.b and 6.c). The obtained results were quite accurate except along the singular line in the direction of the parallax of the reference point. The singular line is evident in Fig. 6.f.

The singularities can be removed and the quality of the computed structure can be improved either by using multiple frames (when their epipoles are non-colinear) or by using multiple reference points (An additional reference point should be chosen so that it does *not* lie on the singular line of the first reference point, and should be first verified to move consistently with the first reference point through the rigidity constraint (5) over a few frames). Of course, combinations of multiple reference points over multiple frames can also be used. Fig. 6.g shows an example of recovering structure of an entire scene from *three* frames relative to the same *single* reference point as in Fig. 6.f. The singular line in Fig. 6.f has disappeared.

The ability of obtain relatively good structure information even with respect to a *single* point has several important virtues: (i) Like [23, 3], it does not require the estimation of the epipole, and therefore does not require dense parallax information. (ii) Unlike previous methods for recovering structure, (including [23, 3]), it provides the capability to handle dynamic scenes, as it does not require having a collection of image points which is known *a priori* to belong to the single $3D$ moving object. (iii) Since it relies on a single parallax vector, it provides a natural continuous way to bridge the gap between $2D$ cases, that assume only planar motion exists, and $3D$ cases that rely on having parallax data.

4.3 New View Generation

The parallax rigidity constraint (5) can be used for generating novel views using a set of "model" views, without requiring any epipolar geometry or shape estimation. This work is still in early stages, and therefore no experimental results are provided.

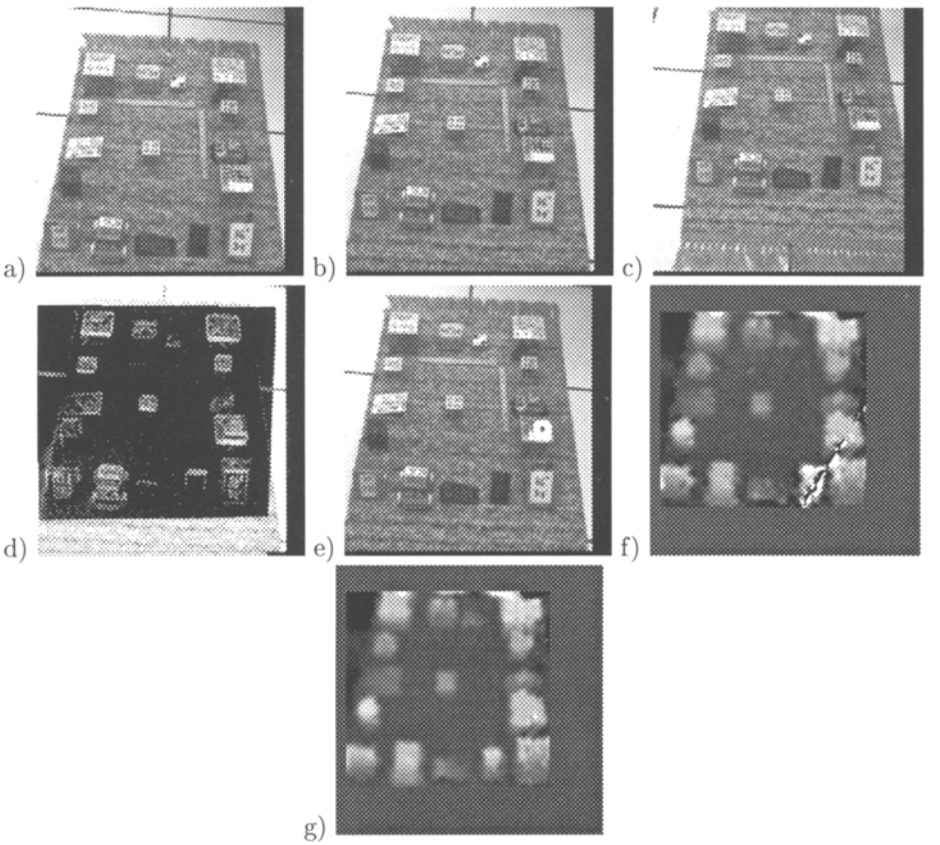

Fig. 6. Shape recovery relying on a single parallax vector.
(a,b,c) Three image frames from a sequence obtained by a hand-held camera of a rug covered with toy cars and boxes. The middle frame (Fig. 6.b) was chosen as the frame of reference. (d) Differences taken after 2D image registration. The detected 2D planar motion was that of the rug, and is canceled by the 2D registration. All other scene parts (i.e., toys and boxes) are misregistered. (e) The selected point of reference (a point on one of the boxes in the bottom right) highlighted by a white circle. (f) The recovered relative structure of the entire scene from two *frames (Figs. 6.b and 6.c) relative to the selected point of reference. Regions close to the image boundary were ignored. The interpreted relative heights were quite accurate except along the singular line in the direction of the parallax displacement of the reference point. (g) The recovered relative structure of the entire scene using all three frames with respect to the selected point of reference. Regions close to the image boundary were ignored. The singular line has disappeared, providing more accurate shape.*

Methods for generating new views based on recovering epipolar geometry (e.g., [4]) are likely to be more noise sensitive than methods that generate the new view based on $2D$ information alone [16], i.e., without going from $2D$ through a $3D$ medium in order to reproject the information once again onto a new $2D$ image plane (the virtual view).

Given two "model" frames, planar parallax motion can be computed for all image points between the first (reference) frame and the second frame. An image point with non-zero parallax is selected, and a "virtual" parallax vector is defined for that point from the reference frame to the "virtual" frame to be generated. The rigidity constraint (Eq. 5) then specifies a single constraint on the virtual parallax motion of all other points from the reference frame to the *virtual* frame. Since each $2D$ parallax vector has two components (i.e., two unknowns), at least two "virtual" parallax vectors are needed to be specified in order to solve for all other virtual parallax vectors. Once the virtual parallax vectors are computed, the new virtual view can be created by warping the reference image twice: First, warping each image point by its computed virtual parallax. Then, globally warping the entire frame with a $2D$ virtual planar motion for the virtual homography.

Note that two virtual parallax vectors may not provide sufficient constraints for some image points. This is due to unfavorable location of those points in the image plane with respect to the two selected reference points and their parallax vectors. However, other image points, for whom the constraint was robust and sufficient to produce reliable virtual parallax, can be used (once their virtual parallax has been computed) as additional points to reliably constrain the virtual parallax of the singular points.

5 Conclusion

This paper presented geometric relationships between the image motion of *pairs* of points over multiple frames. This relationship is based on the *parallax* displacements of points with respect to an arbitrary planar surface, and does not involve epipolar geometry. We derived constraint over two frames relating the projective structure (with respect to the plane) of any pair of points, based only on their image coordinates and their parallax motion. We also derived a $3D$-rigidity constraint between pairs of points over multiple frames.

We showed applications of these parallax-based constraints to the recovery of $3D$ scene structure, to the detection of moving objects in the presence of camera induced motion, and to "new view generation". Our approach can handle difficult situations for $3D$ scene analysis, e.g., where there is only a small set of parallax vectors, and in the presence of independently moving objects. The use of the parallax constraints derived here provides a continuum between "$2D$ algorithms" and the "$3D$ algorithms" for each of the problems mentioned above.

Finally, we outlined the generalization of our parallax based constraints to full image motion (as opposed to *parallax* motion), even when the homography is unknown. This is useful for handling scenes that do not contain a physical planar surface.

References

1. J.R. Bergen, P. Anandan, K.J. Hanna, and R. Hingorani. Hierarchical model-based motion estimation. In *ECCV*, pages 237–252, Santa Margarita Ligure, May 1992.
2. J.R. Bergen, P.J. Burt, R. Hingorani, and S. Peleg. A three-frame algorithm for estimating two-component image motion. *PAMI*, 14:886–895, September 1992.
3. S. Carlsson. Duality of reconstruction and positioning from projective views. In *Workshop on Representations of Visual Scenes*, 1995.
4. O. Faugeras and L. Robert. What can two images tell us about a third one? In *ECCV*, pages 485–492, May 1994.
5. O.D. Faugeras. What can be seen in three dimensions with an uncalibrated stereo rig? In *ECCV*, pages 563–578, Santa Margarita Ligure, May 1992.
6. O.D. Faugeras and B. Mourrain. On the geometry and algebra of the point and line correspondences between n images. In *ICCV*, pages 951–956, June 1995.
7. M. Irani and P. Anandan. Parallax geometry of pairs of points for 3d scene analysis. Technical report, David Sarnoff Research Center, October 1995.
8. M. Irani, B. Rousso, and S. Peleg. Computing occluding and transparent motions. *IJCV*, 12(1):5–16, January 1994.
9. M. Irani, B. Rousso, and S. Peleg. Recovery of ego-motion using image stabilization. In *CVPR*, pages 454–460, Seattle, Wa., June 1994.
10. J.J. Koenderink and A.J. van Doorn. Representation of local geometry in the visual system. *Biol. Cybern.*, 55:367 - 375, 1987.
11. R. Kumar, P. Anandan, and K. Hanna. Shape recovery from multiple views: a parallax based approach. In *DARPA IU Workshop*, November 1994.
12. J.M. Lawn and R. Cipolla. Robust egomotion estimation from affine motion parallax. In *ECCV*, pages 205–210, May 1994.
13. F. Meyer and P. Bouthemy. Region-based tracking in image sequences. In *ECCV*, pages 476–484, Santa Margarita Ligure, May 1992.
14. T. Chang R. Hartley, R. Gupta. Stereo from uncalibrated cameras. In *CVPR-92*.
15. Harpreet Sawhney. 3d geometry from planar parallax. In *CVPR*, June 1994.
16. A. Shashua. Algebraic functions for recognition. *PAMI*, 17:779–789, 1995.
17. A. Shashua and N. Navab. Relative affine structure: Theory and application to 3d reconstruction from perspective views. In *CVPR*, pages 483–489, 1994.
18. M. E. Spetsakis and J. Aloimonos. A unified theory of structure from motion. In *DARPA IU Workshop*, pages 271–283, 1990.
19. W.B. Thompson and T.C. Pong. Detecting moving objects. *IJCV*, 4:29–57, 1990.
20. P.H.S. Torr and D.W. Murray. Stochastic motion clustering. In *ECCV*, 1994.
21. P.H.S. Torr, A. Zisserman, and S.J. Maybank. Robust detection of degenerate configurations for the fundamental matrix. In *ICCV*, pages 1037–1042, 1995.
22. P.H.S. Torr, A. Zisserman, and D.W. Murray. Motion clustering using the trilinear constraint over three views. In *Workshop on Geometric Modeling and Invariants for Computer Vision*, 1995.
23. D. Weinshall, M.Werman, and A. Shashua. Shape descriptors: Bilinear, trilinear and quadlinear relations for multi-point geometry, and linear projective reconstruction algorithms. In *Workshop on Representations of Visual Scenes*, 1995.
24. Andrew Zisserman. A case against epipolar geometry. In *Applications of Invariance in Computer Vision*, pages 35–50, Ponta Delgada, Azores, October 1993.